ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI BANYAKNYA KLAIM ASURANSI KENDARAAAN BERMOTOR MENGGUNAKAN MODEL REGRESI *ZERO-INFLATED* POISSON
(Studi Kasus di PT. Asuransi Sinar Mas Cabang Semarang Tahun 2010)

Muhammad Taufan[1], Suparti[2], Agus Rusgiyono[2]
[1]Alumni Jurusan Statistika FSM UNDIP
[2]Staf Jurusan Statistika FSM UNDIP

Abstract

Poisson regression is one of model that is often used to model the relationship between response variables in the form of discrete data with a set of predictor variables in the form of continuous, discrete, category, or mixture data. In Poisson regression assumes that the mean of the response variable equal to the variance (equidispersion). But in reality, sometimes found a condition called overdispersion, that the variance value is greater than the mean. One of the cause of overdispersion is excess zero in the response variable. One of model that can be used to overcome this overdispersion problem is Zero-Inflated Poisson (ZIP) regression model. This model is applied on a case study of motor vehicle insurance in the branch of PT. Asuransi Sinar Mas in Semarang in 2010 to determine the effect of age of car and types of coverage to number of claims filed by the policyholder to the branch of PT. Asuransi Sinar Mas in Semarang. In this case, the occurrence of zeros due to many policyholders did not file a claim to the branch of PT. Asuransi Sinar Mas in Semarang. From the analytical result obtained the conclution that the age of car and types of coverage affect number of claims filed by the policyholder to the branch of PT. Asuransi Sinar Mas in Semarang in 2010.

Keywords: Poisson Regression, Overdispersion, Zero-Inflated Poisson (ZIP) Regression

1. Pendahuluan

Dewasa ini industri asuransi di tanah air tumbuh cukup menggembirakan. Sejalan dengan pertumbuhan ekonomi nasional yang semakin baik, kinerja industri ini juga makin baik. Dalam lima tahun terakhir industri asuransi tumbuh rata-rata di angka 20%. Sejak tahun 2005, kinerja keuangan industri asuransi nasional memberikan *trend* positif dengan pertumbuhan yang cukup tinggi. Setiap tahun pertumbuhan rata-rata premi bruto asuransi sebesar 23%. Begitu pula pertumbuhan aset sekitar 25 % setiap tahun[9]. Menurut data Badan Pengawas Pasar Modal dan Lembaga Keuangan (Bapepam-LK) tahun 2010, aset industri asuransi pada semester pertama 2010 mengalami peningkatan 18.5% pada asuransi umum dan peningkatan 23.37% pada asuransi jiwa dibandingkan dengan periode yang sama tahun 2009. Besarnya peluang pangsa pasar dan potensi pemasaran asuransi di Indonesia, membuat jumlah perusahaan asuransi meningkat. Industri perasuransian tahun 2010 diisi oleh 367 perusahaan perasuransian yang terdiri dari 142 perusahaan asuransi dan perusahaan reasuransi serta 234 perusahaan penunjang usaha asuransi[2].

Dengan semakin banyaknya perusahaan asuransi, akan terjadi persaingan di antara mereka. Mereka berlomba-lomba memasarkan produknya kepada masyarakat. Berbagai layanan dan ragam produk ditingkatkan demi mendapatkan konsumen. Hal ini juga terjadi pada perusahaan asuransi kendaraan bermotor[2]. Salah satu layanan yang ditawarkan adalah tentang proses pengajuan klaim. Sekarang proses pengajuan klaim asuransi lebih mudah dibandingkan beberapa waktu yang lalu, sehingga para konsumen tidak merasa sungkan dalam mengajukan klaim kepada pihak asuransi[11].

Berawal dari keadaan tersebut, dalam tulisan ini dibahas hasil penelitian tentang analisis faktor-faktor yang mempengaruhi banyaknya klaim yang diajukan pemegang polis kepada pihak perusahaan asuransi. Salah satu metode untuk menganalisis faktor-faktor tersebut adalah dengan menggunakan analisis regresi. Analisis regresi digunakan untuk menganalisis data variabel respon yang berupa data diskrit maupun kontinu. Karena banyaknya klaim asuransi merupakan variabel respon yang berupa data diskrit maka salah satu model regresi yang dapat digunakan adalah model regresi Poisson[3].

Regresi Poisson adalah suatu metode statistika yang digunakan untuk melakukan analisis terhadap data diskrit (*count data*) yang menyatakan banyaknya suatu kejadian yang jarang terjadi dalam suatu era dan selang waktu tertentu. Suatu ciri dari distribusi Poisson adalah adanya equidispersi, yakni keadaan dimana nilai mean dan varians dari variabel respon bernilai sama. Namun pada prakteknya, kadang-kadang ditemukan suatu keadaan yang disebut overdispersi, yakni nilai variansnya lebih besar dari nilai meannya. Salah satu penyebab terjadinya overdispersi adalah adanya terlalu banyak nilai nol (*excess zero*) pada variabel respon. Salah satu metode yang digunakan dalam mengatasi masalah overdispersi akibat adanya terlalu banyak nol (*excess zero*) pada variabel respon adalah metode regresi *Zero-Inflated* Poisson (ZIP)[3].

Dalam penelitian ini data yang digunakan adalah data tentang asuransi mobil tahun 2010 yang diperoleh dari PT. Asuransi Sinar Mas Cabang Semarang yang meliputi data banyaknya klaim yang diajukan pemegang polis kepada pihak PT. Asuransi Sinar Mas Cabang Semarang, umur mobil dan jenis pertanggungan. Sehingga rumusan masalah dalam tulisan ini adalah menentukan model regresi *Zero-Inflated* Poisson (ZIP) dalam mengatasi masalah overdispersi akibat *excess zero* pada regresi Poisson untuk mengetahui pengaruh umur mobil dan jenis pertanggungan terhadap banyaknya klaim yang diajukan pemegang polis kepada pihak PT. Asuransi Sinar Mas Cabang Semarang tahun 2010. Adapun jenis pertanggungan asuransi meliputi *All Risk, Total Lost Only* (TLO) serta gabungan antara All Risk dan *Total Lost Only* (TLO). Tujuan dari penelitian ini untuk menganalisis faktor-faktor yang mempengaruhi banyaknya klaim yang diajukan pemegang polis kepada pihak PT. Asuransi Sinar Mas Cabang Semarang tahun 2010 menggunakan pemodelan regresi *Zero-Inflated* Poisson (ZIP).

2. Tinjauan Pustaka

2.1 *Generalized Linear Model* (GLM)

Generalized Linear Model (GLM) merupakan perluasan dari proses pemodelan linier untuk pemodelan data yang mengikuti distribusi probabilitas selain distribusi normal, seperti Poisson, Binomial, multinomial, dan lain-lain.

Generalized Linear Model didefinisikan ke dalam tiga komponen[10], yaitu:

1. Komponen Acak
   Variabel respon $Y = (y_1, y_2, ..., y_n)$ dalam suatu n observasi diasumsikan saling bebas dan memiliki distribusi yang termasuk dalam keluarga eksponensial, dengan fungsi kepadatan peluang sebagai berikut:
   $$f(y_i;\theta_i,\phi) = \exp\left\{\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i;\phi)\right]\right\} \quad (1)$$
   Parameter θ_i disebut dengan parameter natural. Parameter ϕ disebut dengan parameter dispersi.

2. Komponen Sistematis
   Misalkan $x_{i1},...,x_{ip}$ menunjukkan nilai dari sejumlah p variabel prediktor untuk observasi ke-i. Komponen sistematis yang merupakan komponen kedua dalam GLM, menghubungkan parameter $\{\eta_i\}$ dengan variabel prediktor menggunakan prediktor linier.
$$\eta_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}$$
$$\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_{j1} x_{ij} \qquad (2)$$
   dengan i = 1, 2, …, n dan j = 1, 2, ..., p
   Atau dalam bentuk matriks dapat dinyatakan dengan
$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$
   dengan η adalah vektor (n x 1) dari observasi, X adalah matriks (n x c) dari variabel prediktor [X = $(x_1, x_2, \ldots x_n)^T$], dan β adalah matriks (c x 1) dari koefisien regresi, dengan c = $p+1$.
3. Fungsi *Link*
   Fungsi link adalah suatu fungsi yang menghubungkan komponen acak dengan komponen sistematis. Diketahui $\mu_i = E(y_i)$. Model yang menghubungkan μ_i dengan prediktor linier η_i, dinyatakan dengan
$$g(\mu_i) = \eta_i = \sum_j \beta_j x_{ij}$$
   dengan fungsi g menunjukkan fungsi *link*. Suatu fungsi *link* disebut fungsi *link* kanonik jika $g(\mu_i) = \theta_i$
$$\theta_i = \sum_j \beta_j x_{ij}$$
   dengan θ_i merupakan parameter kanonik.

2.2 Regresi Poisson

Model regresi Poisson adalah model regresi nonlinier yang berasal dari distribusi Poisson yang biasanya digunakan untuk menganalisis data dengan respon berupa variabel diskrit yang nilainya berupa integer tidak negatif.

Misalnya y_i, i = 1,2,… merupakan jumlah kejadian yang muncul dalam selang waktu dengan rata-rata μ_i. Jika Y adalah variabel acak Poisson dengan parameter $\mu > 0$, maka fungsi massa peluangnya adalah :
$$f(y;\mu) = \frac{e^{-\mu}\mu^y}{y!} \quad , y = 0,1,2,\ldots \qquad (3)$$
dengan E(y) = Var(y) = μ

Persamaan (3) dapat ditulis dalam bentuk
$$f(y;\mu) = exp[y_i\, ln(\mu_i) - \mu_i - ln(y_i!)]$$
$$= exp[y_i\theta_i - b(\theta_i) - ln(y_i!)]$$
Persamaan tersebut merupakan suatu bentuk persamaan fungsi distribusi keluarga eksponensial pada persamaan (1), dengan $\theta_i = ln(\mu_i)$, $b(\theta_i) = \mu_i = exp(\theta_i)$, $a(\phi) = 1$, $c(y_i,\phi) = -\,ln(y_i!)$, $E(y_i) = b'(\theta_i) = exp(\theta_i) = \mu_i$ dan $Var(y_i) = b''(\theta_i)a(\phi) = exp(\theta_i) = \mu_i$

Berdasarkan konsep GLM untuk distribusi Poisson bahwa pada saat $g(\mu_i)$ sama dengan parameter natural θ_i (($g(\mu_i) = \theta_i = \ln(\mu_i)$)), sehingga kanonikal link (fungsi yang mentransformasikan nilai mean ke parameter natural) adalah log natural link $g(\mu_i) = \ln(\mu_i)$. Sehingga hubungan μ_i dengan prediktor linier η_i dinyatakan dengan

$\ln(\mu_i) = \eta_i$. Dengan menggunakan fungsi link log natural tersebut diperoleh model regresi Poisson dalam bentuk:

$$\begin{aligned} \ln \mu_i &= \eta_i \\ \ln \mu_i &= \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} \\ \mu_i &= \exp(\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}) \end{aligned} \quad (4)$$

2.2.1 Estimasi Parameter Regresi Poisson

Penaksiran koefisien parameter regresi Poisson menggunakan metode maksimum likelihood yaitu dengan melakukan turunan parsial fungsi ln-likelihood terhadap parameter yang akan diestimasi dan diiterasikan dengan menggunakan metode *Iterative Reweighted Least Square* (IRWLS). Adapun fungsi ln-likelihood untuk regresi Poisson adalah

$$ln\, L(\, y;\beta\,) = \sum_{i=1}^{n} y_i \, ln\, \mu_i - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} ln\, y_i! \quad (5)$$

Estimasi awal dari β dapat diperoleh dengan menggunakan metode *Ordinary Least Square* (OLS) yakni

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y \quad (6)$$

Untuk estimasi selanjutnya dari β digunakan metode *Iterative Reweighted Least Square* (IRWLS). Diperoleh persamaan sebagai berikut:

$$\mathbf{b}^{(m+1)} = (\mathbf{X}^T\mathbf{W}^{(m)}\mathbf{X})^{-1}\, \mathbf{X}^T\mathbf{W}^{(m)}\mathbf{z}^{(m)} \quad (7)$$

2.2.2 Pengujian Kesesuaian Model Regresi Poisson

Untuk menguji kesesuaian / ketepatan model regresi Poisson dapat digunakan statistik uji rasio likelihood dengan prosedur pengujian seperti berikut:

Hipotesis:

H_0 : $\beta_1 = \beta_2 = \ldots = \beta_p = 0$

H_1 : paling sedikit terdapat satu $\beta_j \neq 0$, dengan $j = 1, 2, \ldots, p$

Statistik uji:

$$G = -2\ln\left(\frac{L_0}{L_1}\right) \sim \chi^2_{(p)}$$

dengan L_0 adalah likelihood tanpa variabel bebas dan L_1 adalah likelihood dengan variabel bebas.

Kriteria Uji: H_0 ditolak jika statistik uji $G > \chi^2_{(\alpha;p)}$

2.2.3 Pengujian Signifikansi Parameter Regresi Poisson secara Individu

Jika uji kesesuaian model menunjukkan bahwa model regresi Poisson dapat digunakan maka selanjutnya dilakukan uji signifikansi terhadap masing-masing koefisien regresi dengan uji Wald dengan prosedur pengujiannya sebagai berikut

Hipotesis:

H_0 : $\beta_j = 0$

H_1 : $\beta_j \neq 0$, untuk suatu $j = 1, 2, \ldots, p$

Statistik uji: $W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \sim \chi^2_{(1)}$

Kriteria Uji: H_0 ditolak jika statistik uji $W_j > \chi^2_{\alpha;1}$

2.2.4 Overdispersi

Suatu ciri dari distribusi Poisson adalah adanya equidispersi, yakni keadaan dimana nilai mean dan varians dari variabel respon bernilai sama. Namun pada prakteknya, kadang-kadang ditemukan suatu keadaan yang disebut overdispersi, yakni nilai variansnya lebih besar dari nilai meannya[3].

Untuk menguji asumsi equidispersi pada regresi Poisson dilakukan dengan melihat nilai statistik *Pearson's chi square* yang dibagi dengan derajat bebasnya (*n-p-1*).

$$\hat{\phi} = \frac{\chi^2}{n-p-1} \text{ dengan } \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\text{var}(\hat{\mu}_i)}$$

Apabila nilai taksiran dispersi sama dengan satu, maka asumsi equidispersi terpenuhi. Data dikatakan mengalami overdispersi apabila nilai taksiran dispersi lebih dari satu. Sedangkan data dikatakan mengalami underdispersi apabila nilai taksiran dispersi kurang dari satu[1].

2.3 Regresi *Zero-Inflated* Poisson (ZIP)

Salah satu penyebab terjadinya overdispersi adalah lebih banyak observasi bernilai nol daripada yang ditaksir untuk model Regresi Poisson. Salah satu metode analisis yang diusulkan untuk lebih banyak observasi bernilai nol daripada yang ditaksir adalah model regresi *Zero-Inflated* Poisson (ZIP)[7].

Jika y_i adalah variabel random independen yang mempunyai distribusi ZIP, nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Keadaan pertama disebut *zero state* terjadi dengan probabilitas ω_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut Poisson *state* terjadi dengan probabilitas $(1-\omega_i)$ dan berdistribusi Poisson dengan mean μ_i[7]. Proses dua keadaan ini memberikan distribusi campuran dua komponen dengan fungsi probabilitas sebagai berikut :

$$P(Y = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu_i}, & \text{untuk } y_i = 0 \\ \dfrac{(1-\omega_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!}, & \text{untuk } y_i > 0,\ 0 \le \omega_i \le 1 \end{cases} \quad (8)$$

Lambert (1992) menyarankan model gabungan untuk μ dan ω[8],yakni :

$$\ln(\mu) = X\beta \text{ dan logit}(\omega) = \ln\left[\frac{\omega}{1-\omega}\right] = X\gamma \quad (9)$$

2.3.1 Estimasi Parameter Regresi *Zero-Inflated* Poisson

Estimasi parameter regresi ZIP dapat dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode ini biasanya digunakan untuk menaksir parameter suatu model yang diketahui fungsi densitasnya.

Dari persamaan (9) didapat:

$$\mu_i = exp(x_i^T\beta), \omega_i = \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}} \text{ dan } (1-\omega_i) = \frac{1}{1+e^{x_i^T\gamma}} \quad (10)$$

Selanjutnya persamaan (10) disubstitusikan ke persamaan (8) didapat:

$$P(Y = y_i) = \begin{cases} \dfrac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}} + \dfrac{1}{1+e^{x_i^T\gamma}}(\exp(-e^{x_i^T\beta})), & \text{untuk } y_i = 0 \\ \\ \dfrac{\dfrac{1}{1+e^{x_i^T\gamma}}(\exp(-e^{x_i^T\beta}))(e^{x_i^T\beta})^{y_i}}{y_i!}, & \text{untuk } y_i > 0 \end{cases} \tag{11}$$

Selanjutnya dari persamaan (11) dibuat persamaan ln-likelihood sehingga diperoleh:
Untuk $y_i = 0$, fungsi ln $L(\beta,\gamma \mid y_i)$ yakni:

$$\sum_{i=1}^{n} \ln(e^{x_i^T\gamma} + \exp(-e^{x_i^T\beta})) - \sum_{i=1}^{n} \ln(1+e^{x_i^T\gamma})$$

Untuk $y_i > 0$, fungsi ln $L(\beta,\gamma \mid y_i)$ yakni:

$$\sum_{i=1}^{n} (x_i^T\beta) y_i - e^{x_i^T\beta}) - \sum_{i=1}^{n} \ln(1+e^{x_i^T\gamma}) - \sum_{i=1}^{n} \ln y_i! \tag{12}$$

Penjumlahan fungsi ln-likelihood pada persamaan (12) akan menyulitkan perhitungan karena tidak diketahui nilai nol mana yang berasal dari *zero state* dan mana yang berasal dari Poisson *state*, sehingga fungsi ln-likelihood ini tidak dapat diselesaikan dengan metode numerik biasa.

Untuk memaksimalkan fungsi ln-likelihood (12) digunakan algoritma EM (*Expectation Maximization*) yang merupakan salah satu metode optimasi yang banyak digunakan sebagai alternatif dalam memaksimumkan fungsi Likelihood yang mengandung data hilang (*missing*)[6]. Algoritma ini pertama kali diperkenalkan oleh Dempster, Laird, dan Rubin(1977)[5].

Misalkan variabel Y berkaitan dengan variabel indikator Z yaitu:

$$Z = \begin{cases} 1, \text{ jika } y_i \text{ berasal dari } \textit{zero state} \\ 0, \text{ jika } y_i \text{ berasal dari } \textit{Poisson state} \end{cases}$$

Permasalahannya adalah jika nilai variabel respon y_i = 1,2,3,…, maka nilai z_i = 0. Sedangkan jika nilai variabel respon y_i = 0, maka nilai z_i mungkin 0 mungkin 1. Oleh karena itu, nilai z_i dianggap hilang.

Untuk mengatasi hal ini dilakukan estimasi parameter dengan algoritma EM. Langkah-langkahnya adalah sebagai berikut:

1. Menentukan distribusi dari variabel Z
   $P(z_i = 1) = \omega_i$
   $P(z_i = 0) = 1- \omega_i$
   Sehingga z_i ~ Binomial $(1, \omega_i)$, $E(z_i) = \omega_i$ dan $var(z_i) = \omega_i(1- \omega_i)$
2. Membentuk distribusi gabungan antara y_i dan z_i yaitu

$$\begin{aligned} f(y_i, z_i \mid \omega_i, \mu_i) &= f(z_i) f(y_i \mid z_i) \\ &= f(z_i \mid 1, \omega_i) f(y_i \mid z_i, \mu_i) \\ &= (1-\omega_i)^{(1-z_i)} (\omega_i)^{z_i} \left(\frac{\exp(-\mu_i)\mu_i^{y_I}}{y_i!} \right)^{(1-z_i)} \end{aligned} \tag{13}$$

   Kemudian substitusikan persamaan (9) ke persamaan (13) hingga didapat persamaan ln-likelihood:

$$\ln L(\beta,\gamma \mid y,z) = \sum_{i=1}^{n}\left[z_i x_i^T \gamma - \ln\left(1+\exp(x_i^T\gamma)\right)\right]$$
$$+\sum_{i=1}^{n}(1-z_i)(y_i x_i^T \beta - \exp(x_i^T\beta)) - \sum_{i=1}^{n}(1-z_i)\ln y_i! \qquad (14)$$

Persamaan (14) biasa disebut *complete data likelihood*. Persamaan ini yang akan dimaksimumkan menggunakan algoritma EM, dimana parameter β dan γ dapat diestimasi secara terpisah, dengan menuliskan persamaan (14) menjadi:

$$\ln L(\beta,\gamma,y,z) = \ln L(\beta,y,z) + \ln L(\gamma,y,z) - \sum_{i=1}^{n}(1-z_i)\ln(y_i!)$$

dengan

$$\ln L(\gamma,y,z) = \sum_{i=1}^{n}\left[z_i x_i^T \gamma - \ln\left(1+\exp(x_i^T\gamma)\right)\right] \qquad (15)$$

dan

$$\ln L(\beta,y,z) = \sum_{i=1}^{n}(1-z_i)(y_i x_i^T \beta - \exp(x_i^T\beta)) \qquad (16)$$

3. Tahap ekspektasi
   Ganti variabel z_i dengan $z_i^{(k)}$ yang merupakan ekspektasi dari z_i

   $$z_i^{(k)} = E(z_i \mid y_i, \gamma^{(k)}, \beta^{(k)}) \qquad = P(z_i = 1 \mid y_i, \gamma^{(k)}, \beta^{(k)})$$

   Untuk $y_i = 0$, $z_i^{(k)} = \dfrac{1}{1+\exp(-x_i^T\gamma^{(k)} - \exp(x_i^T\beta^{(k)}))}$

   Untuk $y_i > 0$, $z_i^{(k)} = 0$

   Sehingga persamaaan (15) dan (16) menjadi

   $$\ln L(\gamma^{(k)},y,z^{(k)}) = \sum_{i=1}^{n}\left[z_i^{(k)} x_i^T \gamma^{(k)} - \ln\left(1+\exp(x_i^T\gamma^{(k)})\right)\right] \qquad (17)$$

   $$\ln L(\beta^{(k)},y,z^{(k)}) = \sum_{i=1}^{n}(1-z_i^{(k)})(y_i x_i^T \beta^{(k)} - \exp(x_i^T\beta^{(k)})) \qquad (18)$$

4. Tahap maksimalisasi
   Memaksimalkan β dan γ pada persamaan (17) dan (18) dengan menghitung $\beta^{(k+1)}$ dan $\gamma^{(k+1)}$ dengan metode Newton-Raphson.

   Misalkan $\beta^{(k)}$ dan $\gamma^{(k)}$ adalah aproksimasi metode maksimum likelihood untuk mengestimasi $\hat{\beta}$ dan $\hat{\gamma}$. Dengan menggunakan metode Newton Rhapson maka :

   $$\beta^{(k+1)} = \beta^{(k)} - \left(H^{(k)}\right)^{-1} U^{(k)} \qquad (19)$$

   dan

   $$\gamma^{(k+1)} = \gamma^{(k)} - \left(H^{(k)}\right)^{-1} U^{(k)} \qquad (20)$$

   dimana H adalah turunan kedua dari $\ln L(\beta^{(k)},y,z^{(k)})$ dan $\ln L(\gamma^{(k)},y,z^{(k)})$, U adalah turunan pertama dari $\ln L(\beta^{(k)},y,z^{(k)})$ dan $\ln L(\gamma^{(k)},y,z^{(k)})$.

5. Ganti $\beta^{(k)}$ dan $\gamma^{(k)}$ dengan $\beta^{(k+1)}$ dan $\gamma^{(k+1)}$ pada iterasi selanjutnya, kemudian kembali lakukan tahap ekspektasi.
6. Tahap ke-3 dan ke-4 ini dilakukan berulang-ulang sampai diperoleh penaksir parameter yang konvergen ($|\beta^{(k)} - \beta^{(k+1)}| \leq \varepsilon$ dan $|\gamma^{(k)} - \gamma^{(k+1)}| \leq \varepsilon$, biasanya $\varepsilon = 10^{-5}$).

2.3.2 Pengujian Kesesuaian Model Regresi ZIP

Pengujian kesesuaian model regresi ZIP adalah dengan menggunakan Likelihood Ratio (LR) Test dengan prosedur pengujian [8] sebagai berikut:

Hipotesis:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = \gamma_1 = \gamma_2 = \ldots = \gamma_p = 0$

H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$, dengan $j = 1, 2, \ldots, p$

dengan β_j adalah parameter model $\ln(\mu) = X\beta$ ke-j, γ_j adalah parameter model $\text{logit}(\omega) = X\gamma$ ke-j.

Statistika uji: $G = 2 \ln \left[\frac{L(y;\hat{\omega})}{L(y;\hat{\Omega})}\right] \sim \chi^2_{(2p)}$

Kriteria uji: Tolak H_0 pada taraf signifikansi α jika $G_{\text{hitung}} > \chi^2_{\alpha;2p}$

2.3.3 Pengujian Signifikansi Parameter Regresi ZIP secara Individu

Prosedur untuk menguji signifikansi parameter secara individu sebagai berikut:

a. Uji signifikansi parameter model $\ln(\mu) = X\beta$

Hipotesis:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

untuk suatu $j = 1, 2, \ldots, p$

Statistika uji: $W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \sim \chi^2_{(1)}$

Kriteria uji: Tolak H_0 pada taraf signifikansi α jika $W_j > \chi^2_{\alpha;1}$

b. Uji signifikansi parameter model $\text{logit}(\omega) = X\gamma$

Hipotesis:

$H_0 : \gamma_j = 0$

$H_1 : \gamma_j \neq 0$

untuk suatu $j = 1, 2, \ldots, p$

Statistika uji: $W_j = \left(\frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)}\right)^2 \sim \chi^2_{(1)}$

Kriteria uji: Tolak H_0 pada taraf signifikansi α jika $W_j > \chi^2_{\alpha;1}$

3. Metodologi Penelitian

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini berupa data sekunder tentang asuransi mobil yang diperoleh dari PT. Asuransi Sinar Mas Cabang Semarang. Data yang diambil adalah data asuransi mobil dari bulan Januari sampai bulan Desember tahun 2010. Data yang digunakan berupa data banyaknya klaim yang diajukan pemegang polis kepada pihak PT. Asuransi Sinar Mas Cabang Semarang, umur mobil, dan jenis pertanggungan asuransi. Jenis pertanggungan asuransi terdiri dari *All Risk*, *Total Lost Only* (TLO), serta gabungan antara *All Risk* dan *Total Lost Only* (TLO).

3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini berupa variabel respon (Y) yakni banyaknya klaim yang diajukan kepada pihak PT. Asuransi Sinar Mas Cabang Semarang dan variabel prediktor (X) yang meliputi:

1. umur mobil (X_1).
2. jenis pertanggungan asuransi, merupakan variabel *dummy* dengan tiga kategori. Sehingga variabel *dummy* yang dapat dibentuk adalah sebanyak dua kategori, yaitu:

$$X_2 = \begin{cases} 1, \text{ untuk jenis pertanggungan } \textit{Total Lost Only} \\ 0, \text{ untuk jenis pertanggungan lainnya} \end{cases}$$

$$X_3 = \begin{cases} 1, \text{ untuk jenis pertanggungan gabungan } \textit{All Risk} \text{ dan } \textit{Total Lost Only} \\ 0, \text{ untuk jenis pertanggungan lainnya} \end{cases}$$

4. Hasil dan Pembahasan

4.1 Analisis Deskriptif Data

Deskripsi data banyaknya klaim asuransi mobil yang diajukan kepada pihak PT. Asuransi Sinar Mas Cabang Semarang selama tahun 2010 tertera dalam Tabel 1.

Tabel 1. Deskrisi Data Penelitian

	Banyaknya Klaim	Umur Mobil	Jenis Pertanggungan	
			Total Lost Only (TLO)	Gab. All Risk dan Total Lost Only (TLO)
Total Observasi	406	406	406	406
Mean	0.30	4.73	0.54	0.25
Standar Deviasi	0.607	2.683	0.499	0.431
Nilai Minimum	0	1	0	0
Nilai Maksimum	3	12	1	1

Data asuransi mobil yang digunakan dalam penelitian ini sebanyak 406 data. Adapun rata-rata dan standar deviasi dari banyaknya klaim yang diajukan kepada pihak PT. Asuransi Sinar Mas Cabang Semarang untuk polis yang terdaftar pada bulan Januari sampai Desember 2010 adalah masing-masing sebesar 0.30 dan 0.607, yang berarti dari 100 kejadian terdapat 30 kejadian pengajuan klaim, dengan banyaknya klaim berjumlah maksimal tiga kali. Umur mobil yang diasuransikan berkisar antara satu hingga dua belas tahun dengan rata-rata umur mobil berkisar empat tahun dengan standar deviasi sebesar 2.683.

4.2 Pemodelan Regresi Poisson

Berdasarkan hasil pengolahan dengan software SAS 9.0, diperoleh estimasi parameter model regresi Poisson seperti terlihat pada Tabel 2.

Dari Tabel 2, diperoleh model regresi Poisson:

$\ln(\mu) = -0.6189 + 0.0014\, X_1 - 1.5286\, X_2 - 0.0798 X_3$ atau

$\mu = \exp(-0.6189 + 0.0014\, X_1 - 1.5286\, X_2 - 0.0798 X_3)$

Setelah didapatkan model regresi Poisson, selanjutkan dilakukan uji asumsi equidispersi pada regresi Poisson. Apabila asumsi equidispersi terpenuhi, maka dapat dilakukan pengujian model regresi Poisson, tetapi jika terjadi overdispersi, maka dilakukan pemodelan kembali dengan menggunakan metode regresi *Zero-Inflated* Poisson (ZIP).

Tabel 2. Estimasi Parameter Regresi Poisson

Parameter	DF	Estimate	SE	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	- 0.6189	0.2046	- 1.0198	- 0.2179	9.15	0.0025
x_1	1	0.0014	0.0396	- 0.0763	0.0790	0.00	0.9727
x_2	1	- 1.5286	0.2566	- 2.0315	- 1.0257	35.49	< .0001
x_3	1	- 0.0798	0.2053	- 0.4823	0.3226	0.15	0.6974
Scale	0	1.0000	0.0000	1.0000	1.0000		

4.3 Pengujian Asumsi Equidispersi pada Regresi Poisson

Untuk menguji asumsi equidispersi pada model regresi Poisson dilakukan uji dengan cara membagi nilai *Pearson chi-square* dengan derajat bebasnya. Jika nilai *Pearson Chi-Square* dibagi dengan derajat bebasnya sama dengan satu, maka asumsi equidispersi terpenuhi. Apabila nilainya lebih besar dari satu, maka diindikasikan terjadi overdispersi. Sedangkan apabila nilainya kurang dari satu, maka diindikasikan terjadi underdispersi[1]. Berdasarkan pengolahan dengan software SAS 9.0 diperoleh tabel berikut:

Tabel 3. Pearson Chi-Square Model Regresi Poisson

Pearson χ^2	DF	Pearson χ^2/DF
419.1976	402	1.0428

Dari Tabel 3, dapat diketahui bahwa nilai *Pearson Chi-Square* model dibagi derajat bebasnya adalah 1.0428. Nilai tersebut lebih besar dari satu yang berarti dapat disimpulkan bahwa data mengalami overdispersi.

4.4 Pemodelan Regresi *Zero-Inflated* Poisson

Berdasarkan pengolahan dengan software SAS 9.0, diperoleh estimasi parameter pada model regresi *Zero-Inflated* Poisson (ZIP) sebagai berikut:

Tabel 4. Estimasi Parameter Regresi ZIP

Parameter	Estimate	SE	DF	t Value	Pr |t|	Alpha	Lower	Upper
a_0	2.0024	0.6435	406	3.11	0.0020	0.05	0.7374	3.2673
a_1	- 0.6896	0.1991	406	- 3.46	0.0006	0.05	- 1.0809	- 0.2983
a_2	1.4520	0.8539	406	1.70	0.0898	0.05	- 0.2265	3.1306
a_3	- 1.7910	0.8972	406	- 2.00	0.0466	0.05	- 3.5548	- 0.0271
b_0	0.6420	0.3052	406	2.10	0.0360	0.05	0.0421	1.2419
b_1	- 0.1750	0.0587	406	- 2.98	0.0031	0.05	- 0.2904	- 0.0595
b_2	- 1.0034	0.4186	406	- 2.4	0.0170	0.05	- 1.8264	- 0.1804
b_3	- 0.5049	0.2683	406	- 1.88	0.0606	0.05	- 1.0323	0.0225

Berdasarkan Tabel 4, maka model $\ln(\mu) = X\beta$ yakni :

$$\ln(\mu) = 0.6420 - 0.1750\ X_1 - 1.0034\ X_2 - 0.5049\ X_3$$

Model $\text{logit}(\omega) = X\gamma$ yakni :

$$\text{logit}(\omega) = 2.0024 - 0.6896\ X_1 - 1.4520\ X_2 - 1.7910\ X_3$$

4.5 Kesesuaian Model Regresi *Zero-Inflated* Poisson

Uji kesesuaian model digunakan untuk mengetahui kesesuaian atau kecocokan model regresi *Zero-Inflated* Poisson. Berdasarkan hasil pengolahan dengan software SAS 9.0 diperoleh:

Tabel 5. Nilai *Fits Statistics* Regresi ZIP

-2 Log Likelihood	501.3
AIC (smaller is better)	517.3
AICC (smaller is better)	517.7
BIC (smaller is better)	549.4

Dari Tabel 5 dapat diketahui bahwa nilai G hitung = 501.3. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;6)}$ = 12.59. Karena nilai G hitung > $\chi^2_{(0.05;6)}$, maka H_0 ditolak. Sehingga model regresi ZIP dapat digunakan untuk mengetahui pengaruh umur mobil dan jenis pertanggungan terhadap banyaknya klaim yang diajukan pemegang polis kepada pihak PT. Asuransi Sinar Mas Cabang Semarang tahun 2010.

4.6 Signifikansi Parameter Regresi *Zero-Inflated* Poisson

a. Model ln(μ) = Xβ

Berdasarkan Tabel 4 diperoleh hasil analisis:

1. Pengaruh X_1 (umur mobil) terhadap Y (banyaknya klaim)
   Nilai t_1 hitung = -2.98 sehingga nilai W_1 = 8.8804. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;1)}$ = 3.841. Karena nilai $W_1 > \chi^2_{(0.05;1)}$, maka H_0 ditolak. Sehingga umur mobil berpengaruh terhadap banyaknya klaim.
2. Pengaruh X_2 (jenis pertanggungan TLO) terhadap Y (banyaknya klaim)
   Nilai t_2 hitung = -2.40 sehingga nilai W_2 = 5.76. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;1)}$ = 3.841. Karena nilai $W_2 > \chi^2_{(0.05;1)}$, maka H_0 ditolak. Sehingga jenis pertanggungan TLO berpengaruh terhadap banyaknya klaim.
3. Pengaruh X_3 (jenis pertanggungan gabungan antara *All Risk* dan TLO) terhadap Y (banyaknya klaim)
   Nilai t_3 hitung = -1.88 sehingga nilai W_3 = 3.5344. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;1)}$ = 3.841. Karena nilai $W_3 < \chi^2_{(0.05;1)}$, maka H_0 diterima. Sehingga jenis pertanggungan gabungan antara *All Risk* dan TLO tidak berpengaruh terhadap banyaknya klaim.

b. Model logit(ω) = Xγ

Berdasarkan tabel 4 diperoleh hasil analisis:

1. Pengaruh X_1 (umur mobil) terhadap Y (banyaknya klaim)
   Nilai t_1 hitung = -3.46 sehingga nilai W_1 = 11.9716. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;1)}$ = 3.841. Karena nilai $W_1 > \chi^2_{(0.05;1)}$, maka H_0 ditolak. Sehingga umur mobil berpengaruh terhadap banyaknya klaim.
2. Pengaruh X_2 (jenis pertanggungan TLO) terhadap Y (banyaknya klaim)
   Nilai t_2 hitung = 1.70 sehingga nilai W_2 = 2.89. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;1)}$ = 3.841. Karena nilai $W_2 < \chi^2_{(0.05;1)}$, maka H_0 diterima. Sehingga jenis pertanggungan TLO tidak berpengaruh terhadap banyaknya klaim.
3. Pengaruh X_3 (jenis pertanggungan gabungan antara *All Risk* dan TLO) terhadap Y (banyaknya klaim)
   Nilai t_3 hitung = -2.00 sehingga nilai W_3 = 4.00. Sedangkan dari tabel χ^2, nilai $\chi^2_{(0.05;1)}$ = 3.841. Karena nilai $W_3 > \chi^2_{(0.05;1)}$, maka H_0 ditolak. Sehingga jenis pertanggungan gabungan antara *All Risk* dan TLO berpengaruh terhadap banyaknya klaim.

4.7 Interpretasi Model

Berdasarkan uji signifikansi parameter regresi ZIP diperoleh model:

$$\ln(\mu) = 0.6420 – 0.1750\ X_1 – 1.0034\ X_2 - 0.5049\ X_3$$

dan

$$\text{logit}(\omega) = 2.0024 – 0.6896\ X_1 – 1.4520\ X_2 – 1.7910\ X_3$$

Artinya bahwa:

1. Setiap perubahan satu tahun dalam umur mobil menyebabkan penurunan nilai harapan banyaknya klaim sebesar exp(-0.1750) kali.
2. Setiap pemilihan jenis pertanggungan TLO menyebabkan penurunan nilai harapan banyaknya klaim sebesar exp(-1.0034) kali jenis pertanggungan *All Risk*.
3. Setiap pemilihan jenis pertanggungan gabungan TLO dan *All Risk* menyebabkan penurunan nilai harapan banyaknya klaim sebesar exp(-0.5049) kali jenis pertanggunggan *All Risk*.
4. Setiap perubahan satu tahun dalam umur mobil menyebabkan penurunan resiko tidak mengajukan klaim sebesar exp(-0.6896) kali.
5. Setiap pemilihan jenis pertanggungan . TLO menyebabkan penurunan resiko tidak mengajukan klaim sebesar exp(-1.4520) kali. jenis pertanggunggan *All Risk*
6. Setiap pemilihan jenis pertanggungan gabungan *All Risk* dan TLO menyebabkan penurunan resiko tidak mengajukan klaim sebesar exp(-1.7910) kali jenis pertanggunggan *All Risk*

5. Kesimpulan

Berdasarkan hasil analisis terhadap model regresi Zero Inflated Poisson (ZIP) dapat disimpulkan bahwa umur mobil dan jenis pertanggungan *Total Lost Only* (TLO) serta pertanggungan gabungan All Risk dan *Total Lost Only* (TLO) berpengaruh terhadap banyaknya pengajuan klaim asuransi di PT. Asuransi Sinar Mas Cabang Semarang tahun 2010. Adapun model regresi ZIP yang dihasilkan adalah:

$$\ln(\mu) = 0.6420 – 0.1750\ X_1 – 1.0034\ X_2 - 0.5049\ X_3$$

dan

$$\text{logit}(\omega) = 2.0024 – 0.6896\ X_1 – 1.4520\ X_2 – 1.7910\ X_3$$

dengan:

X_1 : umur mobil

X_2 : jenis pertanggungan *Total Lost Only* (TLO)

X_3 : jenis pertanggungan gabungan All Risk dan *Total Lost Only* (TLO)

Dari model regresi ZIP diperoleh hasil bahwa setiap perubahan satu unit dalam umur mobil menyebabkan penurunan nilai harapan banyaknya klaim sebesar exp(-0.1750) kali dan penurunan resiko tidak mengajukan klaim sebesar exp(-0.6896) kali. Setiap pemilihan jenis pertanggungan TLO menyebabkan penurunan nilai harapan banyaknya klaim sebesar exp(-1.0034) kali dan penurunan resiko tidak mengajukan klaim sebesar exp(-1.4520) kali jenis pertanggunggan *All Risk*. Sedangkan setiap pemilihan jenis pertanggungan gabungan *All Risk* dan TLO menyebabkan penurunan nilai harapan banyaknya klaim sebesar exp(-0.5049) kali dan penurunan resiko tidak mengajukan klaim sebesar exp (-1.7910) kali jenis pertanggunggan *All Risk*.

DAFTAR PUSTAKA

1. Agresti, A., *Categorical Data Analysis*, Second Edition, John Wiley and Sons, Inc., New York, 2002.
2. Bapepam-LK,. *2010 Annual Report*, Bapepam-LK, Jakarta, 2010

3. Cameron, A.C., and Trivedi, P.K., *Regression Analysis of Count Data*, Cambridge University Press,New York, 1998.
4. DeMaris, A., *Regression with Social Data : Modeling Continuous and Limited Response Variables*, John Wiley and Sons, Inc., New Jersey, 2004.
5. Dempster, A., Laird, N.M., and Rubin, D.B, Maximum Likelihood from Incomplete Data via EM Algorithm, *Journal of The Royal Statistical Society*, Series B (Methodological), 1977, Vol. 39, No.1: 1-38.
6. Hall, D.B., and Shen, J., Robust Estimation for Zero-Inflated Poisson Regression, *Scandinavian Journal of Statistics*, Blackwell Publishing Ltd., 2009 : 1-16.
7. Jansakul, N. and Hinde, J.P., Score Tests for Zero-Inflated Models, *Computational Statistics and Data Analysis*, 2002, Vol.40 : 75-96.
8. Lambert, D., Zero-Inflated Poison Regression with an Application to Defects in Manufacturing, *Technometrics*, 1992, Vol.34 : 1-14.
9. Marketeers, *Bagaimana Prospek Industri Asuransi pada Tahun 2011*, 2010, URL: http://the-marketeers.com/archives/bagaimana-prospek-industri-asuransi-pada-tahun-2011.html, diakses 28 Juni 2011.
10. McCullagh, P., and Nelder, J.A., *Generalized Linear Models*, Second Edition, Chapman and Hall, London, 1989.
11. Surabaya Post, *Industri Asuransi Tepis Stigma Persulit Klaim*, 2011, URL: http://www.surabayapost.co.id, diakses 20 Oktober 2011.